УДК 004.932

АНОНИМИЗАЦИЯ ОБЪЕКТОВ НА ФОТОГРАФИЯХ КАК МЕТОД ЗАЩИТЫ ПЕРСОНАЛЬНЫХ ДАННЫХ ПРИ АННОТИРОВАНИИ ДАННЫХ ДЛЯ МАШИННОГО ОБУЧЕНИЯ

Третьяк И.Н.

магистрант,

Институт сферы обслуживания и предпринимательства (филиал) ДГТУ в г. Шахты,

Шахты, Россия

Кузнецов Д.В.

магистрант,

Институт сферы обслуживания и предпринимательства (филиал) ДГТУ в г. Шахты,

Шахты, Россия

Попов А.Э.

к.т.н., доцент,

Институт сферы обслуживания и предпринимательства (филиал) ДГТУ в г. Шахты,

Шахты, Россия

Аннотация

Данная статья посвящена исследованию основных методов анонимизации объектов на фотографиях при аннотировании данных для машинного обучения в контексте защиты персональных данных (ПД). В статье кратко рассмотрена нормативно-правовая база защиты ПД, а также обобщено понятие анонимизации данных. Методы анонимизации объектов на фотографиях

рассмотрены на примере анонимизации лиц. В качестве основного метода анонимизации объектов при разработке клиент-серверного приложения для аннотирования данных предлагается использовать размытие по Гауссу.

Ключевые слова: защита персональных данных, анонимизация данных, анонимизация объектов, детекция объектов, детекция лиц, размытие по Гауссу.

ANONYMIZATION OF OBJECTS IN PHOTOS AS A METHOD OF PERSONAL DATA PROTECTION DURING DATA ANNOTATING FOR MACHINE LEARNING

Tretiak I.N.
master's student,
Institute of Service and Business (branch) DSTU in Shakhty,
Shakhty, Russia

Kuznetsov D.V.
master's student,
Institute of Service and Business (branch) DSTU in Shakhty,
Shakhty, Russia

Popov A.E.
PhD, associate professor,
Institute of Service and Business (branch) DSTU in Shakhty,
Shakhty, Russia

Abstract

This paper is devoted to the study of the main methods of anonymization of objects in photos during data annotating for machine learning in the context of personal data

(PD) protection. The paper briefly discusses the regulatory framework for PD protection and summarizes the concept of data anonymization. Methods of anonymization of objects in photos are considered on the example of face anonymization. Gaussian blurring is proposed as the main method of object anonymization when developing a client-server application for data annotation.

Key words: personal data protection, data anonymization, object anonymization, object detection, face detection, Gaussian blurring.

Введение

В эпоху цифровизации и быстрого развития технологий машинного обучения, вопросы защиты персональных данных (ПД) становятся все более актуальными. Сбор больших массивов данных для обучения моделей глубокого и машинного обучения неизбежно влечет за собой риск нарушения приватности и конфиденциальности информации. Это может произойти как в сфере хранения ПД, так и при аннотировании обучающих данных на основе ПД.

Актуальность проблемы защиты ПД обусловлена не только быстрыми темпами развития технологий и увеличением объемов собираемых данных, но и усилением законодательных требований к защите ПД во многих странах мира.

В рамках данной работы исследуется возможность использования эффективных методов анонимизации данных, которые могут сбалансировать потребности в обучении моделей машинного обучения и необходимость обеспечения конфиденциальности и защиты персональных данных при сборе массива данных для машинного обучения.

Предметом исследования данной статьи является проблема защиты персональных данных при сборе массива данных для машинного обучения. Целью исследования является разработка методов и подходов, которые позволят обеспечить надежную защиту персональных данных в процессе сбора и

обработки данных для машинного обучения в рамках разработки клиент-серверного приложения для аннотирования изображений.

Нормативно-правовая база защиты персональных данных

Сбор и обработка персональных данных для машинного обучения представляют собой важную задачу, которая требует строгого соблюдения правил конфиденциальности и приватности. При неправильной обработке персональных данных возникают риски, такие как утечка информации, неправомерное использование данных, нарушение приватности и потенциальные угрозы для прав и свобод личности.

В соответствии с Федеральным законом от 27 июля 2006 г. № 152-ФЗ «О персональных данных» [1], операторы и иные лица, получившие доступ к персональным данным, обязаны не раскрывать третьим лицам и не распространять персональные данные без согласия субъекта персональных данных. Это подчеркивает важность защиты персональных данных при сборе набора данных для машинного обучения.

Но не только в России есть законы, которые ограничивают доступ к персональным данным, к примеру на территории Евросоюза (ЕС) действует Генеральный регламент ЕС о защите персональных данных (GDPR) [2], который усиливает и унифицирует защиту персональных данных всех лиц в Европейском союзе. Он направлен на то, чтобы дать гражданам контроль над собственными персональными данными и упростить нормативную базу для международных экономических отношений.

Таким образом, необходимость защиты персональных данных при сборе набора данных для машинного обучения обусловлена не только законодательными требованиями различных стран и политических объединений, но и этическими принципами, а также потребностью в обеспечении безопасности и защите прав и свобод каждого человека.

Анонимизация данных

Анонимизация данных играет критическую роль в обработке сырых данных для машинного обучения, особенно когда эти данные подлежат последующей разметке с помощью человека [3].

Анонимизация данных – это процесс, при котором персональные данные пользователя изменяются таким образом, что становится невозможным связать данные с конкретным пользователем. Это особенно важно при работе с большими наборами сырых данных, где необходимо обеспечить конфиденциальность пользовательской информации.

Однако, анонимизация данных – это сложный процесс, который требует тщательного подхода. Если анонимизация проводится неправильно, это может привести к значительной потере информации, что, в свою очередь, может снизить качество моделей машинного обучения.

Важно отметить, что анонимизация данных не всегда является полностью безопасной. Например, даже после анонимизации данных, существуют методы деанонимизации, которые могут использоваться для восстановления идентификации пользователя. Поэтому важно использовать современные и надежные методы анонимизации, а также регулярно обновлять их в соответствии с последними научными исследованиями в этой области.

Анонимизация данных является важным инструментом для защиты персональных данных при сборе и последующей разметке сырых данных для машинного обучения. Однако, необходимо тщательно подходить к процессу анонимизации, чтобы обеспечить эффективную защиту данных без значительной потери информации.

Регулирование доступа к данным – это процесс, который обеспечивает, что только авторизованные пользователи могут получить доступ к определенным данным. Это может включать в себя различные механизмы, такие как контроль доступа на основе ролей, аутентификация и авторизация пользователей.

В контексте анонимизации данных, регулирование доступа может быть использовано для обеспечения того, что только авторизованные пользователи могут получить доступ к исходным, неанонимизированным данным. Это может быть особенно важно в случаях, когда данные содержат чувствительную информацию, и требуется обеспечить их конфиденциальность.

Методы анонимизации объектов на фотографиях

При разработке клиент-серверного приложения для аннотирования изображений следует применять такие методы защиты персональных данных, как шифрование данных при передаче по сети, анонимизацию данных, регулирование доступа к данным. Так, обращение к серверу должно осуществляться только по защищенному протоколу HTTPS, а доступ к приложению должен регулироваться при помощи механизмов аутентификации. Для аутентификации пользователей можно применять, например, JWT-токены [4]. Подробнее следует остановиться на механизмах анонимизации.

Рассмотрим методы анонимизации объектов на примере анонимизации лица, так как оно потенциально является объектом, по которому можно идентифицировать человека. При аннотировании изображений может возникнуть такая ситуация, когда на фотографии присутствует лицо человека. Появление лица на фотографии уже можно расценить как утечку ПД, но только в тот момент, когда изображение попадает к другому человеку. До этого момента, при хранении на сервере, фотография с лицом является объектом защиты информации. Следовательно, при передаче фотографии на аннотирование, с лицом человека необходимо произвести некоторые манипуляции.

Следует выделить несколько методов анонимизации лица. Лицо на фотографии можно перекрыть цветной фигурой (прямоугольником или эллипсом), перекрыть цветной фигурой только глаза, а также применить

размытие изображения в границах этих фигур. Также лицо на фотографии можно заменить на сгенерированное несуществующее лицо [5; 6].

Рассмотрим эти методы с точки зрения защиты информации. Так, при сокрытии только глаз, с некоторой вероятностью можно установить личность человека по оставшейся части лица. Поэтому перекрытие лица полностью является более надежным методов. В обоих случаях факт анонимизации не скрывается от разметчика данных. Однако полностью сокрытие лица (как в методе с фигурой) нежелательно, так как это может нарушить визуальное восприятие изображения и, следовательно, снизить качество аннотации. Методы восстановления размытых изображений существуют, но тем вероятность установления личности здесь гораздо ниже, чем при перекрытии только глаз.

Для эффективности метода с заменой лица на сгенерованное необходимо также обеспечить достоверность замены, чтобы при взгляде на изображение у возможных злоумышленников не возникало подозрений о замене лица. К тому же применение данного метода достаточно дорогостоящее с точки зрения потребления вычислительных ресурсов и времени.

Так, при разработке клиент-серверного приложения для аннотирования данных следует использовать метод анонимизации лиц с использованием размытия. В его основе лежит применения методов детекции лиц с последующим использованием алгоритмов размытия изображений в границах найденного лица.

Рассмотрим три основных метода детекции лиц: метод Виолы-Джонса [7], метод гистограммы направленных градиентов [8] и метод с использованием нейронной сети (на примере нейросети MTCNN [9]).

Метод Виолы-Джонса [7] основан на использовании простых признаков, похожих на признаки Хаара. Здесь используется три вида признаков. Значение признака двух прямоугольников — это разница между суммой пикселей в двух

прямоугольных областях. Эти области имеют одинаковый размер и форму, а также являются соседними по горизонтали или вертикали.

Метод гистограммы направленных градиентов (Histogram of Oriented Gradients, HOG) [8] основан на позонном вычисление градиента яркости изображения, последующем разделении изображении на ячейки и вычисления гистограмм в каждой из них и получением дескрипторов. Границы лиц определяются путем классификации дескрипторов по методу опорных векторов (Support Vector Machine).

Каскад нейронных сетей MTCNN (Multi-task Cascaded Convolutional Networks) [9] состоит из трех сверточных нейронный сетей. На каждом из промежуточных этапов происходит уточнение найденных ограничивающих коробок (bounding boxes) и подавление немаксимумов (non-maximum supppresion, NMS) для объединения очень близких коробок. На последнем этапе формируются окончательные ограничивающие коробки и определяются пять лицевых ориентиров (центры глаз, кончик носа, уголки губ) [9, 1500].

Согласно результатам тестирования, проведенного авторами статьи [10] на наборе данных LFW [11], наиболее точным из представленных алгоритмов, согласно основным метрикам [12], является каскад нейронных сетей MTCNN (Таблица 1). Он смог найти все возможные лица в наборе данных. Таким образом, для задач анонимизации лиц нейросетевые алгоритмы подходят наилучшим образом, так как обеспечивают максимальную эффективность детекции. Скорость работы алгоритма в данном случае не так сильно важна, так как лица будут детектироваться заранее, а координаты ограничивающих коробок помещаться в базу данных.

Таблица 1 — Результаты тестирования методов обнаружения лиц [10, 90]

Алгоритм	Точность	Полнота	F1-метрика
Виола-Джонс	1	0,9976	0,9988

HOG	1	0,9956	0,9978
MTCNN	1	1	1

Пример размытия лица на изображении представлен на рисунке 1. Здесь для детекции лица применялась реализация алгоритма HOG из библиотеки face_recognition [13] для языка Python. Непосредственно для размытия применялся алгоритм размытия по Гауссу. Для дополнительной защиты информации при размытии каждого лица (или другого объекта) следует использовать случайное значение радиуса размытия. В контексте решения задачи классификации изображений (image classification) человек не является целевым классом, поэтому размытие лица не помешает верно идентифицировать объект на изображении (класс «сазан» в примере).

Рис. 1 – Пример исходного изображения лица (слева) и анонимизированного (справа) [Создано авторами]

Следует отметить, что размытие может применяться для анонимизации любых объектов, которые могут быть найдены с помощью алгоритмов детекции объектов (object detection). Так, на рисунке 2 приведен пример размытия фигуры человека, найденной с помощью нейронной сети YOLOv5 [14]. В данном случае, размытые фигуры не препятствуют верной идентификации сцены (сцена «парк»), при аннотировании фотографии для решения задачи

распознавания сцен (scene recognition). Однако, при этом нельзя идентифицировать людей на изображении.

Рис. 2 – Пример исходного изображения сцены (слева) и анонимизированного (справа) [Создано авторами]

Заключение

Исследование подчеркнуло важность защиты ПД в эпоху цифровизации и быстрого развития технологий машинного обучения. Был изучен такой подход к защите ПД в контексте машинного обучения, как анонимизация данных. Методы анонимизации объектов на фотографиях рассмотрены на примере

анонимизации лица. Так, одним из эффективных и простых методов анонимизации объектов является размытие.

Для обеспечения эффективной защиты ПД рекомендуется использовать современные и надежные методы анонимизации данных, а также регулярно обновлять их в соответствии с последними научными исследованиями в этой области. Важно также тщательно подходить к процессу анонимизации, чтобы обеспечить эффективную защиту данных без значительной потери информации. Однако, необходимо также учитывать потенциальные угрозы и атаки и принимать соответствующие меры безопасности.

Библиографический список:

1. Федеральный закон от 27.07.2006 N 152-ФЗ (ред. от 06.02.2023) "О персональных данных" // ЗАКОНЫ, КОДЕКСЫ И НОРМАТИВНО-ПРАВОВЫЕ АКТЫ РОССИЙСКОЙ ФЕДЕРАЦИИ URL: https://legalacts.ru/doc/152_FZ-o-personalnyh-dannyh/ (дата обращения: 17.05.2024).
2. Общий регламент по защите данных // Wikipedia URL: https://ru.wikipedia.org/wiki/Общий_регламент_по_защите_данных (дата обращения: 17.05.2024).
3. КАК GOOGLE АНОНИМИЗИРУЕТ ДАННЫЕ // Google. Политика конфиденциальности и условия использования URL: https://policies.google.com/technologies/anonymization?hl=ru (дата обращения: 17.05.2024).
4. JSON Web Token // Wikipedia URL: https://ru.wikipedia.org/wiki/JSON_Web_Token (дата обращения: 17.05.2024).
5. Hukkelås H., Mester R., Lindseth F. Deepprivacy: A generative adversarial network for face anonymization // International symposium on visual computing. – Cham : Springer International Publishing, 2019. – С. 565-578.

6. Barattin S. et al. Attribute-preserving face dataset anonymization via latent code optimization // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. – 2023. – С. 8001-8010.

7. Viola P., Jones M. Rapid object detection using a boosted cascade of simple features // Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001. – Ieee, 2001. – Т. 1. – С. I-I.

8. Dalal N., Triggs B. Histograms of oriented gradients for human detection // 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05). – Ieee, 2005. – Т. 1. – С. 886-893.

9. Zhang K. et al. Joint face detection and alignment using multitask cascaded convolutional networks // IEEE signal processing letters. – 2016. – Т. 23. – №. 10. – С. 1499-1503.

10. Kuznetsov D. V., Beraza A. N., Dmitrienko N.A. Comparative analysis of face detection methods in Python // Тенденции развития науки и образования. – 2023. – №. 93-8. – С. 88-90. – DOI 10.18411/trnio-01-2023-405. – EDN CQSEUA.

11. Huang G. B. et al. Labeled faces in the wild: A database forstudying face recognition in unconstrained environments // Workshop on faces in'Real-Life'Images: detection, alignment, and recognition. – 2008.

12. Кузнецов Д. В. Методы оценки эффективности нейронных сетей для классификации изображений // Дневник науки. – 2023. – № 4(76). – DOI 10.51691/2541-8327_2023_4_21. – EDN OIFLOV.

13. face_recognition // Github URL: https://github.com/ageitgey/face_recognition (дата обращения: 11.05.2024).

14. Ultralytics YOLOv5 Architecture // Ultralytics URL: https://docs.ultralytics.com/yolov5/tutorials/architecture_description/ (дата обращения: 17.05.2024).

Оригинальность 76%